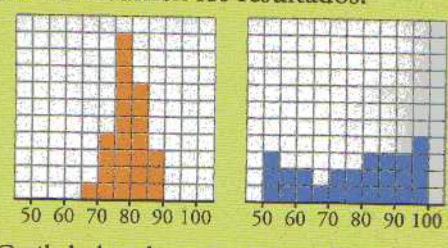
12 Estadística

1 Asigna cada diana, 1, 2, 3, 4, a uno de los tiradores *A*, *B*, *C* o *D*.

2 La máxima puntuación que se puede conseguir en cada tanda de 10 disparos es de 100 puntos.

Dos tiradores con buenas escopetas disparan 25 tandas de 10 disparos. Estas gráficas resumen los resultados.



¿Cuál de los dos es más regular?

1. Soluciones a estos problemas.

En la ilustración tienes las primeras tandas de disparos efectuados por cuatro tiradores.

A es un buen tirador y tiene una buena escopeta.

B es un buen tirador y tiene una escopeta con la mira desviada.

C es un tirador mediocre con una buena escopeta.

D es un tirador mediocre con una escopeta con la mira desviada.

Antes de comenzar, recuerda

Qué es una tabla de frecuencias

Al número de veces que se repite un dato se le denomina **frecuencia** de ese dato.

Una **tabla de frecuencias** es una tabla en la que cada valor de la variable tiene emparejada su frecuencia.

Ejemplo

El número de alumnos y alumnas que han obtenido un 7 en 3.° A es 3.

Lo expresamos así: $f(7) = 3$.

Y se lee así: "frecuencica de 7 es 3".

1 ¿Qué significa el 4 que hay en la columna de la derecha de la tabla de frecuencias?

2 ¿Cuál es la frecuencia de 9?

3 Suma los números de la columna de la derecha. ¿Podrías haber supuesto el resultado sin efectuar la suma?

DISTRIBUCIÓN DE NOTAS DE LOS 36 ALUMNOS Y ALUMNAS DE 3.° A

TABLA DE FRECUENCIAS

VALORES	FRECUENCIA
0	0
1	2
2	4
3	5
4	3
5	8
6	5
7	3
8	1
9	3
10	2

Qué es una gráfica estadística

La representación gráfica de los datos sirve para captar, de un solo golpe de vista, las características más sobresalientes de una distribución de datos.

Ejemplos

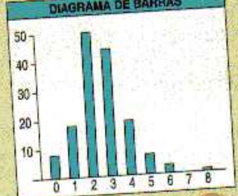
Hay muchos tipos de representaciones gráficas. Las más usuales son:

NÚMERO DE HIJOS QUE TIENEN 150 FAMILIAS

TABLA DE FRECUENCIAS

N.° DE HIJOS	CANTIDAD DE FAMILIAS
0	8
1	18
2	50
3	44
4	19
5	7
6	3
7	0
8	1

DIAGRAMA DE BARRAS

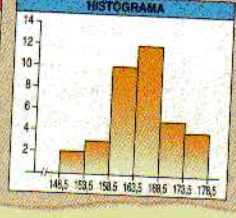


ESTATURAS DE LOS 36 ALUMNOS Y ALUMNAS DE UNA CLASE DE 3.° A

TABLA DE FRECUENCIAS

ESTATURA	N.° DE CHICOS Y CHICAS
148,5 a 153,5	2
153,5 a 158,5	3
158,5 a 163,5	10
163,5 a 168,5	12
168,5 a 173,5	5
173,5 a 178,5	4

HISTOGRAMA



REPARTO, SEGÚN EL TIPO DE TRABAJO, DE 600 000 TRABAJADORES

TABLA DE FRECUENCIAS

SECTOR DE PRODUCCIÓN	N.° DE TRABAJADORES
Agricultura	50 000
Industria	175 000
Servicios	375 000

DIAGRAMA DE SECTORES

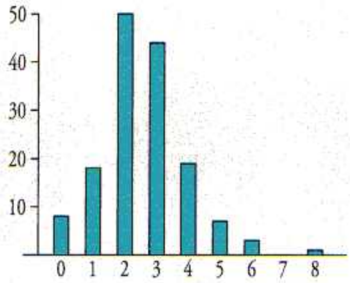
4 Observa el histograma. Si una chica mide 168 cm, ¿en cuál de los intervalos se encuentra? ¿Por qué crees que las barras son anchas, de modo que se juntan unas con otras? ¿Por qué crees que en el diagrama de barras son estrechas y están separadas?

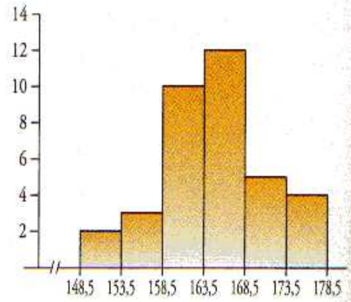
5 Observa la tabla de frecuencias del diagrama de sectores. Calcula el porcentaje de trabajadores que corresponde a cada sector. Reparte los 360° de la circunferencia proporcionalmente a esos porcentajes y obtendrás el ángulo de cada sector.

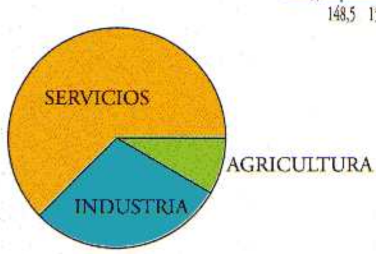
1 Población y muestra

En la página anterior hemos visto tres distribuciones. Cada una de ellas se refiere a un colectivo:

- 150 *familias* de una ciudad.
- Los 36 *alumnos y alumnas* de una clase.
- Los 600 000 *trabajadores* de una cierta comarca.







El colectivo objeto de un estudio estadístico se llama **población.**

A veces, el conjunto que interesa es demasiado numeroso para poder analizar cada uno de sus elementos; entonces se extrae una **muestra.** Por ejemplo, es posible que las 150 familias estudiadas sean una muestra extraída de una población más numerosa: todas las familias de esa ciudad.

De modo que un colectivo es población o muestra según nos interese por sí mismo, o bien sea un medio para inferir información sobre un colectivo más extenso.

Población es el conjunto de todos los elementos objeto de nuestro estudio.

Muestra es un subconjunto, extraído de la población, cuyo estudio sirve para inferir características de toda la población.

Individuo es cada uno de los elementos que forman la población o la muestra.

Por ejemplo: en una gasolinera se pretende hacer un estudio de su clientela. Para ello, se observan y se anotan ciertas características de algunos de los coches que repostan, elegidos al azar.

El conjunto de todos los coches que forman su clientela es la *población.* Los coches seleccionados para ser analizados forman la *muestra.* Cada coche es un *individuo.*

2. Ampliación: El valor de las muestras.

Actividades

1 Un fabricante de tornillos desea hacer un control de calidad. Para ello, recoge 1 de cada 100 tornillos producidos y lo analiza.

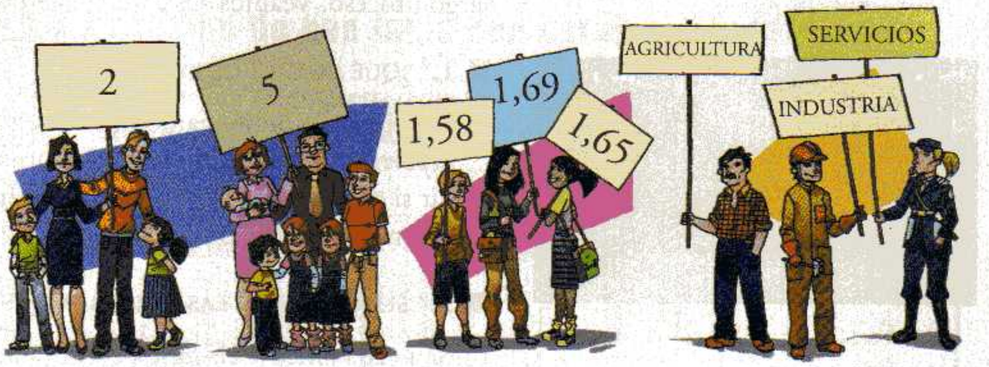
a) ¿Cuál es la población?

b) ¿Cuál es la muestra?

c) ¿Cuáles son los individuos?

2 Variables estadísticas

El *número de hijos*, la *estatura* y el *sector de producción* son las variables que hemos estudiado en las distribuciones anteriores.



NÚMERO DE HIJOS — ESTATURA — SECTOR DE PRODUCCIÓN

Las dos primeras **variables** son **cuantitativas,** porque sus valores se expresan con números (cantidades).

La tercera **variable** es **cualitativa,** porque el sector de producción al que pertenece un trabajador no se expresa mediante un número, sino mediante una cualidad.

Una **variable cuantitativa** es **discreta** cuando solo admite valores aislados (el número de hijos puede ser 2 ó 3, pero no una cantidad intermedia).

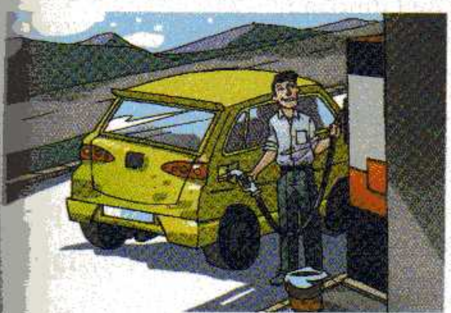
Una **variable cuantitativa** es **continua** cuando entre cada dos valores pueden darse todos los intermedios (una persona puede medir 172,4 cm, aunque habitualmente se redondea y se da un número entero de centímetros).

TIPOS DE VARIABLES ESTADÍSTICAS

Cuantitativa: Numérica.

- **Discreta:** Solo puede tomar valores aislados.
- **Continua:** Podría tomar todos los valores de un intervalo.

Cualitativa: No numérica.



En el ejemplo de la página anterior, supongamos que en cada coche se observa el *número de ocupantes, el tipo de carburante* y el *coste* del producto repostado. Estas tres variables son:

— Número de ocupantes (1, 2, 3, …): cuantitativa discreta.

— Tipo de carburante (gasóleo, súper, …): cualitativa.

— Coste (37,42 €): cuantitativa continua.

Actividades

1 El fabricante de tornillos descrito en la página anterior estudia en cada tornillo si es *correcto* o *defectuoso,* su *longitud* y el *número de pasos de rosca.* Di de qué tipo es cada una de estas variables.

3 El proceso que se sigue en estadística

La información estadística llega a nosotros (a los ciudadanos) mediante gráficas o tablas muy bien construidas, con las que resulta muy sencillo entender la información que se nos da. Sin embargo, esas tablas y gráficas son el resultado de un largo proceso. Veamos sus principales pasos.

1.° ¿QUÉ QUEREMOS ESTUDIAR?

Por ejemplo: Supongamos que deseamos tener información sobre las *aficiones deportivas* de los alumnos y las alumnas de un centro. Antes deberemos decidir si el "deporte" que investigamos interesa como práctica, como espectáculo o como información.

2.° SELECCIÓN DE LAS VARIABLES QUE SE VAN A ANALIZAR

Si a cada chico y chica le preguntamos qué deportes practica, sin más, puede que recibamos una cantidad enorme de posibles respuestas, difíciles de organizar. Y, además, con seguridad, habrá quienes entiendan que practicar un deporte es entregarse a él en alma y vida, y quienes crean que lanzar algún día en el recreo balones a canasta es practicar baloncesto.

Para evitar todos estos inconvenientes, la encuesta debe ser muy clara, con todas las posibles alternativas señaladas. Es decir, *debe ser evidente cuál es la variable y cuáles son sus posibles valores.*

3.° RECOLECCIÓN DE DATOS

Se efectúan las medidas o se realizan las encuestas.

4.° ORGANIZACIÓN DE DATOS

Se ordenan, se pasan a unos papeles o, mejor, se introducen en el ordenador.

PASOS SIGUIENTES

Los pasos siguientes son la elaboración de tablas y gráficas y el cálculo de parámetros. A ellos dedicaremos el resto de esta unidad.

Actividades

1 Se quiere hacer una encuesta para estudiar las aficiones de los jóvenes a la lectura. Di, justificadamente, cuáles de las preguntas siguientes te parecen razonables y cuáles no:

a) Di cuáles son tus lecturas preferidas.

b) De los géneros literarios siguientes, señala aquellos que has leído más de una hora en el último mes:

- ❑ NOVELA ❑ HISTORIA ❑ BIOGRAFÍA
- ❑ POESÍA ❑ TEATRO ❑ FILOSOFÍA

c) ¿Lees periódicos? Si es así, ¿de qué tipo?

d) A cuál o cuáles de las siguientes publicaciones periódicas dedicas más de dos horas semanales:

- ❑ Diarios de actualidad y política.
- ❑ Diarios deportivos.
- ❑ Revistas científicas o de divulgación.
- ❑ Revistas de sociedad (del corazón).
- ❑ Otros. Indica cuáles ....

4 Confección de una tabla de frecuencias

DATOS DESORDENADOS

↓

TABLA CON LOS DATOS ORGANIZADOS

Recuento

Para hacer el recuento, se leen las notas una a una y se hace una señal donde corresponda.

Si las señales se agrupan de cinco en cinco, se cuentan mejor. (La quinta es la horizontal y sirve para cerrar el manojo).

Notación

En las tablas de frecuencias se suele designar:

$x_i \rightarrow$ valores de la variable

$f_i \rightarrow$ valores de la frecuencia

3. Confecciona tablas de frecuencias.

Una vez recogidos los datos, hay que **tabularlos;** es decir, hay que confeccionar una tabla para organizarlos. Esto se consigue con una **tabla de frecuencias.**

Confección de una tabla con datos aislados

Si la variable toma un número reducido de valores, se procede como en el ejemplo siguiente. En él, la variable, x_i, toma los valores 1, 2, 3, …, 10.

NOTAS OBTENIDAS POR UN GRUPO DE ALUMNAS

9	4	8	5	5
4	1	7	2	2
3	9	6	4	10
8	2	1	6	7
6	10	10	8	8
4	6	5	5	10
6	7	2	5	5
3	5	3	6	8

RECUENTO	
1	||
2	||||
3	|||
4	||||
5	𝍸 ||
6	𝍸 |
7	|||
8	𝍸
9	||
10	||||

TABLA DE FRECUENCIAS

x_i	f_i
1	2
2	4
3	3
4	4
5	7
6	6
7	3
8	5
9	2
10	4

Confección de una tabla con datos agrupados en intervalos

Si la variable toma muchos valores, conviene agruparlos en intervalos.

ALTURA DE 30 ALUMNAS Y ALUMNOS DE UNA CLASE

168	160	168	175	168
168	158	149	160	178
158	163	171	162	163
156	154	160	165	165
161	162	166	163	170
164	165	173	172	168

RECUENTO	
Entre 148,5 y 153,5	|
Entre 153,5 y 158,5	||||
Entre 158,5 y 163,5	𝍸 ||||
Entre 163,5 y 168,5	𝍸 𝍸
Entre 168,5 y 173,5	||||
Entre 173,5 y 178,5	||

TABLA RESUMEN

INTERVALO	FRECUENCIA
148,5 a 153,5	1
153,5 a 158,5	4
158,5 a 163,5	9
163,5 a 168,5	10
168,5 a 173,5	4
173,5 a 178,5	2

Actividades

1 Lanzamos dos dados, sumamos las puntuaciones y anotamos los resultados. Repetimos la experiencia 30 veces:

11, 8, 9, 9, 3   4, 11, 7, 7, 8   7, 5, 6, 4, 4

7, 10, 2, 6, 10   7, 7, 6, 2, 8   7, 5, 8, 6, 9

Confecciona una tabla de frecuencias.

2 Con los datos del ejemplo anterior (altura de 30 alumnas y alumnos), efectúa una tabla de frecuencias con los datos agrupados en los intervalos siguientes:

147,5 - 151,5 - 155,5 - 159,5 - 163,5 - 167,5 - 171,5 - 175,5 - 179,5

5 Gráfico adecuado al tipo de información

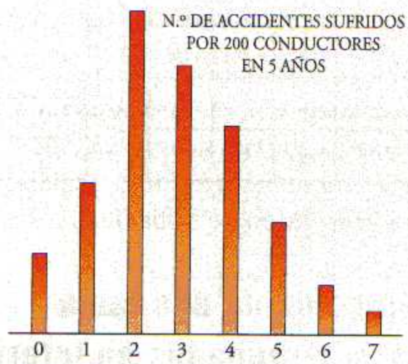
La elaboración de gráficos estadísticos es un arte. En los medios de comunicación encontramos espléndidas representaciones que nos permiten, con un solo golpe de vista, entender de qué se nos habla y asimilar la información que se nos da.

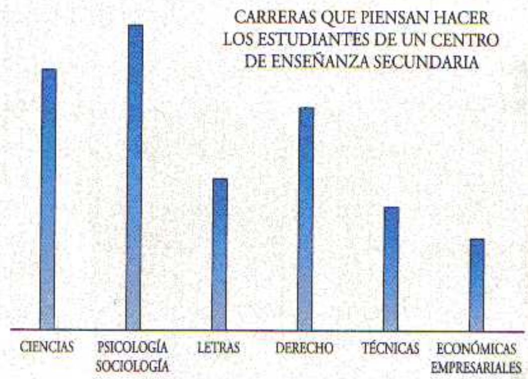
Sin pretender llegar a ese grado de virtuosismo, podemos reflexionar sobre algunas de las claves para utilizar con corrección los tipos de gráficos de uso más frecuente.

Diagrama de barras

El **diagrama de barras** se utiliza para representar tablas de frecuencias correspondientes a **variables cuantitativas discretas.** Por eso, las barras son estrechas y se sitúan sobre los valores puntuales de la variable.

También se utiliza para representar distribuciones de **variables cualitativas.**



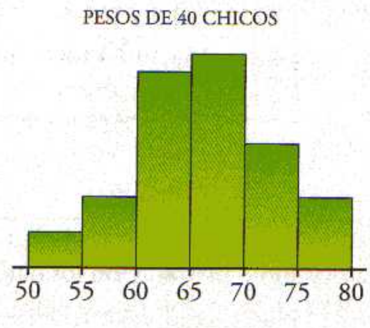


Histograma de frecuencias

El histograma se utiliza para distribuciones de **variable continua.** Por eso se usan rectángulos tan anchos como los intervalos.

Etimología

Histograma: Viene del griego *histos*, que significa "barra" y también "mástil de barco".



Aunque los datos no vengan dados por intervalos (como en el caso de las edades de los niños de una guardería), cuando se trata de una variable continua (1 año significa que aún no ha cumplido 2) es razonable usar el histograma y no el diagrama de barras.

Polígono de frecuencias

El **polígono de frecuencias** se utiliza en los mismos casos que el histograma. Se construye uniendo los puntos medios de los rectángulos y prolongando, al principio y al final, hasta llegar al eje.

Su sentido es suavizar los escalones que se producen en el histograma.

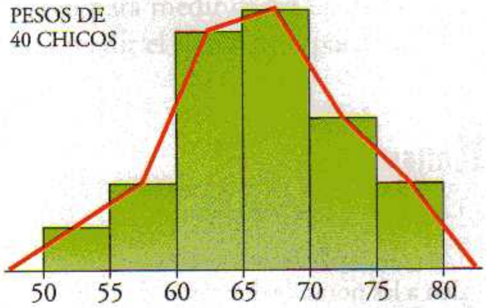
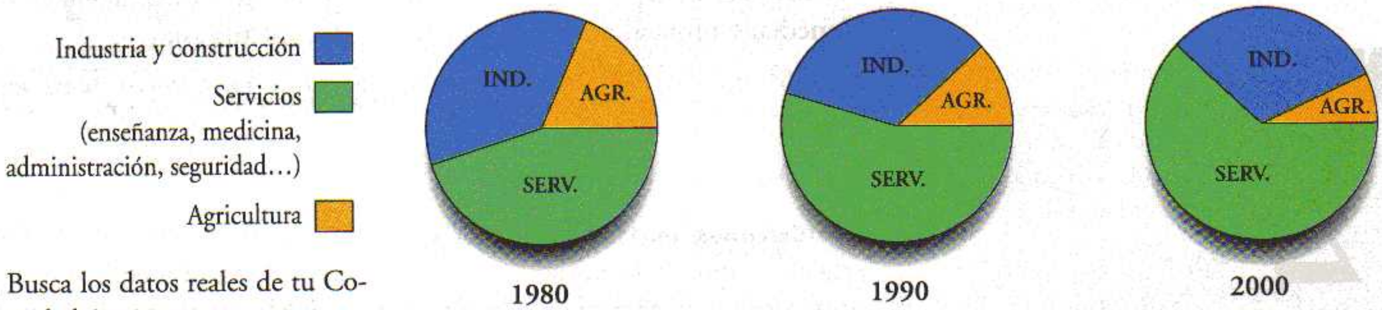


Diagrama de sectores

En un **diagrama de sectores,** el ángulo de cada sector es proporcional a la frecuencia correspondiente.

Se puede utilizar para todo tipo de variables, pero se usa muy frecuentemente para las variables cualitativas.

Este tipo de diagrama es especialmente adecuado para representar, en varios de ellos, diversas situaciones similares y poder establecer comparaciones. Por ejemplo, comparemos el reparto de la población laboral de una cierta comunidad autónoma, según el tipo de trabajo, en los años 1980, 1990 y 2000:



4. Busca los datos reales de tu Comunidad Autónoma.

Actividades

1 Representa, mediante el gráfico adecuado, las tablas estadísticas siguientes:

a) Tiempo que emplean los alumnos y las alumnas de un curso en ir desde su casa al colegio.

TIEMPO (min)	N.° DE ALUMNOS
0 – 5	2
5 – 10	11
10 – 15	13
15 – 20	6
20 – 25	3
25 – 30	1

b) Número de alumnos y alumnas en el curso 2005/06 en una cierta comunidad autónoma, según la etapa de estudios en la que estaban.

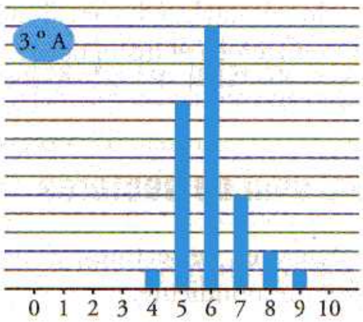
INFANTIL	55 000
PRIMARIA	125 000
SECUNDARIA OBLIGATORIA	100 000
BACHILLERATO Y FORMACIÓN PROFESIONAL	60 000
UNIVERSIDAD	80 000
TOTAL	420 000

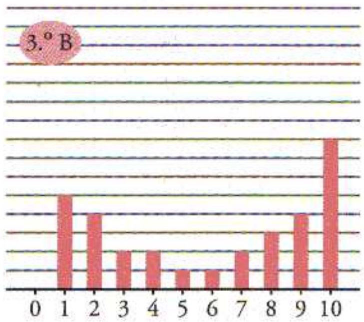
6 Parámetros estadísticos

Los parámetros estadísticos sirven para sintetizar la información dada por una tabla o por una gráfica. Los hay de dos tipos: de **centralización** y de **dispersión.**

Los parámetros de centralización nos indican en torno a qué valor (centro) se distribuyen los datos.

Los parámetros de dispersión nos informan sobre cuánto se alejan del centro los valores de la distribución.





Reflexiona

La media no es suficiente

Las gráficas de la derecha corresponden a las notas de dos clases. En ambas, la nota media es, aproximadamente, 6. La **media** es un parámetro que nos informa sobre el centro alrededor del cual se distribuyen los valores. Pero observa que, aun teniendo la misma media, estas distribuciones son muy distintas. Necesitamos otros parámetros que señalen esas diferencias.

Medidas de centralización

■ **MEDIA**

Si llamamos $x_1, x_2, \ldots, x_n$ a los valores que toma una distribución estadística, la **media,** o promedio, se designa por $\bar{x}$ y se calcula así:

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} \qquad \text{Abreviadamente, } \bar{x} = \frac{\Sigma x_i}{n}$$

■ **MEDIANA**

Si ordenamos los datos de menor a mayor, la **mediana,** *Me,* es el valor que está en medio; es decir, tiene tantos individuos por debajo como por encima.

Si el número de datos fuera par, a la mediana se le asigna el valor medio de los dos términos centrales.

■ **MODA**

La **moda,** *Mo,* es el valor con mayor frecuencia.

Los parámetros media, mediana y moda se llaman **medidas de centralización,** porque alrededor de ellos se distribuyen los valores de la distribución.

Notación

Σ El signo Σ se utiliza para indicar sumas de varios sumandos.

Σx_i se lee:

"suma de los x_i"

Actividades

1 Nos dan la distribución de notas siguiente:

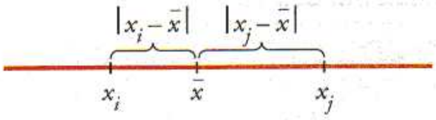
2, 4, 4, 4, 5, 7, 9, 9, 10

a) Comprueba, calculándola, que la nota media es $\bar{x} = 6$.

b) Comprueba que la mediana es $Me = 5$.

c) ¿Cuál es la mediana si suprimimos el 10?

d) ¿Cuál es la moda?



Medidas de dispersión

Vamos a estudiar ahora parámetros que sirven para medir cómo de dispersos están los datos. En todos ellos, la idea clave es medir el grado de separación de los datos a la media.

RECORRIDO O RANGO

Es la diferencia entre el dato mayor y el menor. Es decir, es la longitud del tramo dentro del cual están los datos.

DESVIACIÓN MEDIA

Es el promedio de las distancias de los datos a la media:

$$\text{DM} = \frac{|x_1 - \bar{x}| + |x_2 - \bar{x}| + \ldots + |x_n - \bar{x}|}{n} = \frac{\Sigma\,|x_i - \bar{x}|}{n}$$

VARIANZA

Es el promedio de los cuadrados de las distancias de los datos a la media:

$$\text{Varianza} = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \ldots + (x_n - \bar{x})^2}{n} = \frac{\Sigma\,(x_i - \bar{x})^2}{n}$$

Esta fórmula es equivalente a la siguiente:

$$\text{Varianza} = \frac{x_1^2 + x_2^2 + \ldots + x_n^2}{n} - \bar{x}^2 = \frac{\Sigma x_i^2}{n} - \bar{x}^2$$

DESVIACIÓN TÍPICA, σ

Es la raíz cuadrada de la varianza: $\sigma = \sqrt{\text{varianza}}$

A partir de ahora prestaremos especial atención a los parámetros media ($\bar{x}$) y desviación típica (σ). La información que da cada uno de ellos complementa a la del otro.

Desviación típica

¿Por qué la **desviación típica**?

La varianza tiene un grave inconveniente. Imagina que estamos tratando con una distribución de estaturas dadas en cm. La media vendría dada en cm, pero la varianza vendría en cm^2 (es decir, una superficie en lugar de una longitud). Por eso, extraemos su raíz cuadrada, obteniendo la desviación típica que, en nuestro ejemplo, sí sería una longitud dada en cm.

Ejercicio resuelto

Obtener las medidas de dispersión de la siguiente distribución de notas:

2, 4, 4, 4, 5, 7, 9, 9, 10

RECORRIDO: $10 - 2 = 8$ MEDIA: $\bar{x} = 6$

DESVIACIÓN MEDIA: $\text{DM} = \dfrac{|2-6| + |4-6| + |4-6| + \ldots}{9} = \dfrac{22}{9} = 2{,}44$

VARIANZA: $\text{Var} = \dfrac{(2-6)^2 + (4-6)^2 + (4-6)^2 + \ldots}{9} = \dfrac{64}{9} = 7{,}11$

o bien: $\text{Var} = \dfrac{2^2 + 4^2 + 4^2 + 4^2 + 5^2 + \ldots}{9} - 6^2 = \dfrac{388}{9} - 36 = 7{,}11$

DESVIACIÓN TÍPICA: $\sigma = \sqrt{\text{varianza}} = \sqrt{7{,}11} = 2{,}67$

Actividades

2 Halla las medidas de dispersión de esta distribución de pesos:

83, 65, 75, 72, 70, 80, 75, 90, 68, 72

3 Halla la varianza de la distribución siguiente:

8, 7, 11, 15, 9, 7, 13, 15

Calcúlala utilizando las dos fórmulas de la varianza. Comprueba que es mucho más cómoda la segunda.

7 Cálculo de $\bar{x}$ y σ en tablas de frecuencias

Cuando los datos estadísticos vienen dados en tablas de frecuencias, los cálculos pueden disponerse de forma tal que los parámetros se obtengan con gran comodidad.

x_i	f_i
x_1	f_1
x_2	f_2
…	…
x_n	f_n

■ CÁLCULO DE $\bar{x}$

Veámoslo sobre un ejemplo:

x_i	f_i
4	1
5	10
6	14
7	5
8	2
9	1

Observa la tabla de frecuencias de la izquierda (corresponde a la gráfica de 3.° A, página 256). Para calcular la **media** tendríamos que sumar:

$$4 + \underbrace{(5+5+\ldots+5)}_{10 \text{ veces}} + \underbrace{(6+6+\ldots+6)}_{14 \text{ veces}} + \underbrace{(7+7+\ldots+7)}_{5 \text{ veces}} + 8 + 8 + 9$$

y dividir el resultado por $1 + 10 + 14 + 5 + 2 + 1 = 33$.

Sin embargo, la suma de arriba se obtiene más eficazmente así:

$$4 \cdot 1 + 5 \cdot 10 + 6 \cdot 14 + 7 \cdot 5 + 8 \cdot 2 + 9 \cdot 1$$

Es decir, cada valor de la variable se multiplica por la frecuencia asociada y se suman todos los resultados.

x_i	f_i	$f_i \cdot x_i$
4	1	4
5	10	50
6	14	84
7	5	35
8	2	16
9	1	9
	33	198
	Σf_i	$\Sigma f_i \cdot x_i$

Para facilitar los cálculos, añadimos una nueva columna a la tabla, $f_i \cdot x_i$

El total de individuos se obtiene sumando la columna f_i

$$f_1 + f_2 + \ldots + f_n = 33 \rightarrow \Sigma f_i = 33$$

y la suma de todas las notas se obtiene sumando la columna $f_i\, x_i$

$$f_1 x_1 + f_2 x_2 + \ldots + f_n x_n = 198 \rightarrow \Sigma f_i\, x_i = 198$$

La media es $\bar{x} = \dfrac{\Sigma f_i\, x_i}{\Sigma f_i} = \dfrac{198}{33} = 6.$

Resumimos las conclusiones obtenidas:

En una distribución dada por su tabla de frecuencias, la **media** es:

$$\bar{x} = \frac{\Sigma f_i\, x_i}{\Sigma f_i} \qquad \begin{cases} \Sigma f_i = f_1 + f_2 + \ldots + f_n & \text{(número de individuos)} \\ \Sigma f_i x_i = f_1 x_i + \ldots + f_n x_n & \text{(suma de todos los valores)} \end{cases}$$

Actividades

1 Halla $\bar{x}$ en las distribuciones siguientes:

a) NOTAS (corresponde a la gráfica de 3.° B, página 256):

x_i	0	1	2	3	4	5	6	7	8	9	10
f_i	0	5	4	2	2	1	1	2	3	4	8

b) ESTATURAS:

x_i	151	156	161	166	171	176
f_i	2	5	11	14	5	3

CÁLCULO DE σ

5. Ampliación. Demostración de la equivalencia de las igualdades para la desviación típica.

La **desviación típica** admite dos expresiones equivalentes:

$$\sigma = \sqrt{\frac{\Sigma f_i (x_i - \bar{x})^2}{\Sigma f_i}} \qquad \Sigma f_i (x_i - \bar{x})^2 = f_1 (x_1 - \bar{x})^2 + \dots + f_n (x_n - \bar{x})^2$$

(suma de los cuadrados de las desviaciones a la media)

$$\sigma = \sqrt{\frac{\Sigma f_i x_i^2}{\Sigma f_i} - \bar{x}^2} \qquad \Sigma f_i x_i^2 = f_1 x_1^2 + \dots + f_n x_n^2$$

(suma de los cuadrados de todos los valores)

Con las dos fórmulas se llega al mismo resultado. Sin embargo, es mucho más práctica la segunda de ellas. Veamos por qué:

Puesto que $f_i \cdot x_i^2$ es igual a $(f_i \cdot x_i) \cdot x_i$, añadiendo la columna que se obtiene multiplicando los correspondientes elementos de las columnas 1.ª y 3.ª, se calcula fácilmente la media y la desviación típica. La suma de los elementos de la 2.ª columna, así como los de la 3.ª y los de la 4.ª, nos permite obtener los parámetros buscados:

$(x_i) \cdot (f_i \cdot x_i)$

x_i	f_i	$f_i \cdot x_i$	$f_i \cdot x_i^2$
4	1	4	16
5	10	50	250
6	14	84	504
7	5	35	245
8	2	16	128
9	1	9	81
	33	198	1 224
	Σf_i	$\Sigma f_i \cdot x_i$	$\Sigma f_i \cdot x_i^2$

MEDIA: $\bar{x} = \frac{198}{33} = 6$

DESVIACIÓN TÍPICA:

$$\sigma = \sqrt{\frac{1\,224}{33} - 6^2} = 1{,}04$$

Tablas con datos agrupados en intervalos

Cuando tenemos los datos agrupados en intervalos (en lugar de valores puntuales), se le asigna a cada intervalo su valor central, llamado **marca de clase.** Se obtiene así una tabla de frecuencias como la de arriba. Por ejemplo:

ESTATURA	PERSONAS	MARCAS DE CLASE	x_i	f_i
153,5 a 158,5	5	$\rightarrow \frac{153{,}5 + 158{,}5}{2} = 156 \rightarrow$	156	5
158,5 a 163,5	11	$\rightarrow \frac{158{,}5 + 163{,}5}{2} = 161 \rightarrow$	161	11
163,5 a 168,5	14	$\rightarrow \frac{163{,}5 + 168{,}5}{2} = 166 \rightarrow$	166	14
168,5 a 173,5	5	$\rightarrow \frac{168{,}5 + 173{,}5}{2} = 171 \rightarrow$	171	5

Actividades

2 Halla σ en las distribuciones de la actividad 1 de la página anterior.

8 Obtención de $\bar{x}$ y σ con calculadora

x_i	f_i
151	2
156	4
161	11
166	14
171	5
176	4

Calculadora

Casi todas las calculadoras científicas están preparadas para el cálculo de los parámetros $\bar{x}$ y σ.

Las orientaciones que aquí se ofrecen son generales, ya que cada modelo de calculadora tiene una nomenclatura y unos procedimientos propios. Por tanto, **investiga en tu calculadora** y consulta su manual de instrucciones.

Ayuda

Si en el teclado de tu calculadora no aparecen explícitamente las teclas de resultados:

$$n,\ \Sigma x \text{ y } \Sigma x^2$$

búscalos mediante las secuencias

RCL 3, RCL 2, RCL 1

Calculadoras de pantalla sencilla

Estudiemos con un ejemplo (observa la tabla de la izquierda) los pasos que hay que dar para introducir eficazmente unos datos en la calculadora y conseguir los correspondientes resultados.

PASOS QUE SE DEBEN DAR	EJEMPLO
① **Preparación.** Pon el aparato en disposición de realizar cálculos estadísticos: * MODO SD. Analiza en tu calculadora cómo se consigue este modo	MODE * → [SD]
② **Borra** los datos que puedan haberse quedado acumulados de un trabajo anterior. (En algunas calculadoras, aunque se apaguen, estos datos no se borran).	INV AC
③ **Introduce** los datos. Cada dato se introduce poniéndolo en la pantalla y pulsando la tecla DATA. Si el dato está n veces, se pulsará n veces la tecla DATA, o bien se hará: dato ⊠ n DATA Sigue hasta cargar todos los datos.	151 ⊠ 1 DATA → [151] 156 ⊠ 4 DATA → [156] 161 ⊠ 9 DATA → [161] 166 ⊠ 10 DATA → [166] 171 ⊠ 4 DATA → [171] 176 ⊠ 2 DATA → [176]
④ **Corrige.** Posibilidad de borrar. Si has introducido un dato erróneamente, puedes eliminarlo escribiéndolo en pantalla y pulsando INV DATA.	Dato erróneo: 181 ⊠ 6 DATA Bórralo: 181 ⊠ 6 INV DATA
⑤ **Resultados.** Pulsa las teclas: n → número de individuos → $n = \Sigma f_i$ Σx → suma de todos los valores → $\Sigma x = \Sigma f_i x_i$ Σx^2 → suma de los cuadrados de los valores → $\Sigma x^2 = \Sigma f_i x_i^2$ $\bar{x}$ → media σ_n → desviación típica y obtendrás el valor correspondiente.	n → [30] Σx → [4920] Σx^2 → [807910] $\bar{x}$ → [164] σ_n → [5.859465]

Esta consulta la puedes hacer en cualquier momento del proceso. Después, si lo deseas, puedes seguir introduciendo datos.

Actividades

1 Sigue el proceso anterior para calcular $\bar{x}$ y σ en la distribución de notas de la actividad 1 de la página 258.

2 Utiliza la calculadora para obtener $\bar{x}$ y σ en la distribución de estaturas de la actividad 1 de la página 258.

Atención

Es posible que la tabla que encuentras, después de implantar STAT, 1-VAR, no tenga la columna de frecuencias. En tal caso, teclea esta secuencia

[SHIFT] [SETUP] ▼ 3(STAT)

1(FRECUENCY ON)

x_i	f_i
151	2
156	4
161	11
166	14
171	5
176	4

Calculadoras de pantalla descriptiva

① Preparación para trabajar en estadística

Hemos de encontrar las referencias STAT (estadística) 1-VAR (con una variable). La secuencia puede ser: [MODE] 2 → 1.

Aparece en la pantalla una tabla en la que se irán situando los valores de la variable, x, y sus correspondientes frecuencias, FREQ.

	x	FREQ
1	▓	
2		
3		

② Borrar los datos acumulados del trabajo anterior

Si al encender la calculadora se encuentra preparada para el tratamiento estadístico (STAT en la parte alta de la pantalla) se recupera la tabla mediante la secuencia [SHIFT] [STAT] 2(DATA). Si la tabla contiene datos que no deseamos conservar, se borran volviendo a instalar el tratamiento estadístico ([MODE] 2 → 1).

③ Introducción de datos

Empezamos introduciendo en la tabla todos los valores de la variable:

151 [=] 156 [=] 161 [=] … 176 [=]

	x	FREQ
5	171	1
6	176	1
7	▓	

Observa que asigna, automáticamente, valores 1 en las correspondientes frecuencias.

Para introducir los verdaderos valores de las frecuencias, utilizamos el cursor ▶ ▲▲▲… para situarnos en el lugar correspondiente. Ahora introducimos los valores de f_i:

1 [=] 4 [=] 9 [=] 10 [=] 4 [=] 2 [=]

	x	FREQ
5	171	4
6	176	2
7	▓	▓

Concluye tecleando [AC] para salir de la tabla.

Si queremos volver a ella ponemos [SHIFT] [STAT] 2(DATA).

④ Corregir.

Si hay algún error, con el cursor nos posicionamos en él, tecleamos el valor correcto y pulsamos [=].

⑤ Resultados.

Se acccede a n, Σx, Σx^2, $\bar{x}$ y σ así:

n (n.° de individuos: Σf_i):	[SHIFT] [STAT] 5(VAR)1(n) [=]	→	30
Σx (suma de los valores: $\Sigma f_i x_i$):	[SHIFT] [STAT] 4(SUM)2(Σx) [=]	→	4 920
Σx^2 (suma de los cuadrados: $\Sigma f_i x_i^2$):	[SHIFT] [STAT] 4(SUM)1(Σx^2) [=]	→	807 910
$\bar{x}$ (media):	[SHIFT] [STAT] 5(VAR)2($\bar{x}$) [=]	→	164
σ (desviación típica):	[SHIFT] [STAT] 5(VAR)3($x\sigma n$) [=]	→	5,859465

Actividades

3 Sigue el proceso anterior para calcular $\bar{x}$ y σ en la distribución de notas de la actividad 1 de la página 258.

4 Utiliza la calculadora para obtener $\bar{x}$ y σ en la distribución de estaturas de la actividad 1 de la página 258.

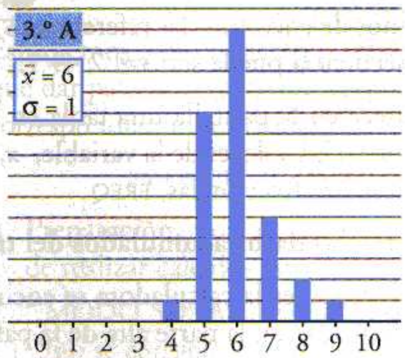
9 Interpretación conjunta de $\bar{x}$ y σ

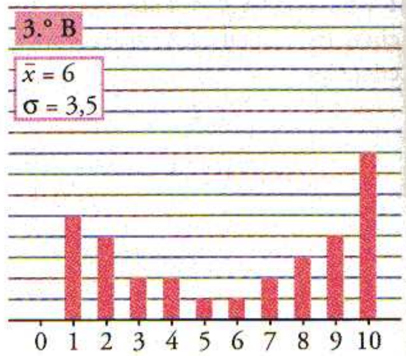
Reflexiona

Interpretación gráfica de σ

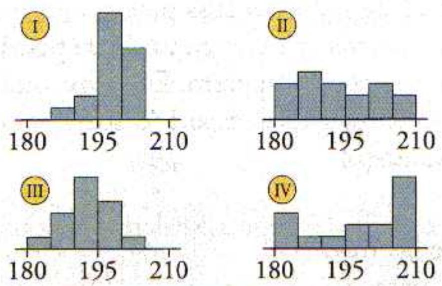
Observa que la desviación típica de la distribución de 3.º B es mucho mayor que la de 3.º A, porque en 3.º B los datos están más alejados de la media: hay muchas notas bajas (1, 2) y muchas altas (8, 9, 10), mientras que en la de 3.º A la mayor parte de las notas son intermedias (5, 6, 7).

Ahora que ya conocemos sus parámetros, mira de nuevo las dos gráficas con las que iniciamos la página 256.





Conociendo los valores de $\bar{x}$ y σ, podemos obtener una idea relativamente buena de cómo es una distribución. La **media** nos dice dónde está su **centro.** La **desviación típica** orienta sobre cómo de alejados de la media, cómo de **dispersos,** están los datos.



EQUIPO	$\bar{x}$	σ
A	198,5	9,7
B	198,1	3,9
C	193	4,6
D	193,4	8,1

Ejercicio resuelto

Las cuatro gráficas del margen corresponden a las estaturas de los jugadores de cuatro equipos de baloncesto, A, B, C y D, cuyos parámetros son los que aparecen más abajo. ¿Cuál es la gráfica de cada equipo?

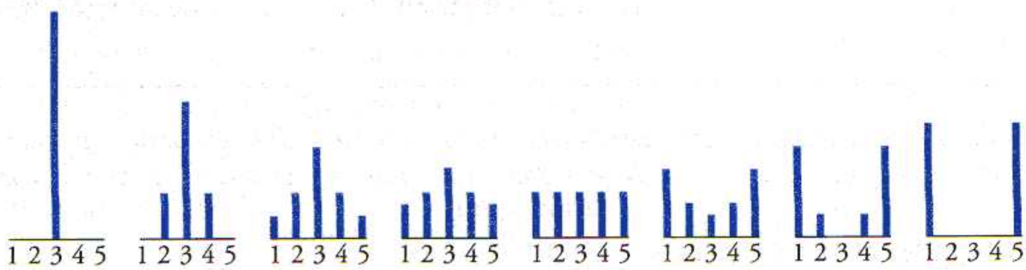
Si nos fijamos en las gráficas, es fácil ver que los equipos I y IV tienen medias superiores a 195, mientras que las medias de II y III son inferiores. Por tanto, podemos asociar así: A y B ↔ I y IV   C y D ↔ II y III

Por otra parte, es claro que los jugadores de IV tienen estaturas más extremas que los de I. Según eso, su desviación típica será mayor, por lo que A ↔ IV, B ↔ I.

Análogamente, C ↔ III y D ↔ II.

Actividades

1 Para entrenarte en la comparación de desviaciones típicas, observa esta secuencia de distribuciones. Todas tienen la misma media, pero sus desviaciones típicas son cada vez mayores.



Para convencerte de que es así, observa que para pasar de cada gráfica a la siguiente hay que acortar las barras centrales y alargar las exteriores. Es decir, hay que alejar de la media algunos individuos.

6. Refuerza. Interpretación conjunta de $\bar{x}$ y σ.

Coeficiente de variación

Los pesos de los toros de lidia de una ganadería se distribuyen con una media $\bar{x} = 500$ kg y una desviación típica $\sigma = 40$ kg.

Los pesos de los perros de una exposición canina tienen una media $\bar{x} = 20$ kg y una desviación típica $\sigma = 10$ kg.

La desviación típica de los pesos de la manada de toros bravos (40 kg) es superior a la de los perros (10 kg). Sin embargo, los 40 kg son poca cosa para el enorme tamaño de los toros (es decir, los toros de esa manada son *muy parecidos* en peso), mientras que 10 kg es mucho en relación con el peso de un perro. En casos como este, la desviación típica no es una medida adecuada para comparar dispersiones. Por ello, definimos un nuevo parámetro estadístico.

	$\bar{x}$	σ
TOROS	500	40
PERROS	20	10

40 con relación a 500 es menor que 10 con relación a 20.

Para comparar la dispersión de dos poblaciones heterogéneas, se define el **coeficiente de variación** así:

$$CV = \frac{\sigma}{\bar{x}}$$

Al dividir σ entre $\bar{x}$ estamos relativizando la dispersión.

El resultado se da, a veces, en tantos por ciento.

En el ejemplo de los toros y los perros, obtenemos:

- Para los toros: $CV = \frac{40}{500} = 0{,}08$ Es decir, el 8%.
- Para los perros: $CV = \frac{10}{20} = 0{,}50$ Es decir, el 50%.

De este modo sí se aprecia claramente que la variación de los pesos de los perros (50%) es mucho mayor que la de los pesos de los toros (8%).

	$\bar{x}$	σ	CV
TOROS	500	40	8%
PERROS	20	10	50%

Actividades

2 En distintas tiendas de instrumentos musicales preguntamos el precio de ciertos modelos concretos de piano, flauta travesera y armónica. Los resultados obtenidos tienen las siguientes medias y desviaciones típicas:

	PIANOS	FLAUTAS	ARMÓNICAS
MEDIA	943 €	132 €	37 €
DESV. TÍPICA	148 €	22 €	12 €

Compara la variación de los precios de estos tres productos.